DOCUMENT RESUME

ED 198 178 TM 810 146

AUTHOR Pfau, Richard H.
TITLE Standardizing Behavioral Measurements Across Cultures and Time.
PUB DATE Feb 81
NOTE 22p.: Paper presented at the Annual Meeting of the Society for Cross-Cultural Research (10th, Syracuse, NY, February 20-22, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Behavior Rating Scales: *Behavior Standards: *Cross Cultural Studies: Research Methodology: Research Needs
IDENTIFIERS *Behavior Observation Systems: *Category Systems: Standardization

ABSTRACT

Behavior can be described and measured in many ways. One approach in particular provides a potential for making precise and valid cross-cultural comparisons of behavior not equaled by other techniques. This approach is one in which low-inference observation instruments called "category systems" are used to measure behaviors observed. This paper: (1) summarizes why category systems appear to be suitable for comparing behaviors cross-culturally in standardized ways: (2) indicates how category systems and variations of these instruments have been used in the past by cross-cultural researchers and points out some accomplishments and limitations of past usage: (3) indicates procedures which can help to ensure that these instruments are used in standardized ways in future studies: and (4) indicates research needed to help investigators make more standardized measurements of behavior in different cultures and at different times. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED198178

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

STANDARDIZING BEHAVIORAL MEASUREMENTS ACROSS CULTURES AND TIME

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Richard H. Pfau
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Richard H. Pfau
IPS International/School of Education
University of Connecticut
Storrs, CT 06268

Abstract

This paper summarizes why category systems appear suitable for comparing behaviors across cultures in standardized ways, indicates how category systems and variations of these systems have been used in the past by cross-cultural researchers, points out accomplishments and limitations of past usage, indicates procedures which can help to ensure that these instruments are used in standardized ways in future studies, and suggests research needed to help investigators make more standardized measurements of behavior in different cultures and at different times.

Paper presented at the 10th annual meeting
of the Society for Cross-Cultural Research
Syracuse, New York, 20-22 February 1981

TM 810146

Behavior can be described and measured in many ways. Relatively indirect ways include the use of questionnaires, interviews, and diaries where persons describe their own or the behavior of others. More direct ways include having a person directly observe behaviors and record their observations by writing narrative descriptions or by using rating systems, checklists, or other instruments. It appears, however, that one approach in particular provides a potential for making precise and valid cross-cultural comparisons of behavior not equaled by other techniques. This approach is one in which low-inference observation instruments called "category systems" are used to measure behaviors observed (Pfau, 1976, 1980).

This paper (a) summarizes why category systems appear to be suitable for comparing behaviors cross-culturally in standardized ways, (b) indicates how category systems and variations of these instruments have been used in the past by cross-cultural researchers and points out some accomplishments and limitations of past usage, (c) indicates procedures which can help to ensure that these instruments are used in standardized ways in future studies, and (d) indicates research needed to help investigators make more standardized measurements of behavior in different cultures and at different times.

Suitability of Category Systems for Comparative Research

Category systems are systematic observation instruments which are characterized by their focus on relatively specific, well-defined categories of behavior, and by the use of counting techniques and other approaches based on time to measure the extent of occurrence of the behaviors being

focused upon.[1] The strength of these instruments for comparative research purposes is derived from their use of low-inference classification and recording procedures.

> Przeworski and Teune have stated that
>
> whether two or more phenomena are "comparable" depends upon whether their properties have been expressed in a standard language. A language of measurement defines classes of phenomena by providing specific criteria for deciding whether an observation can be assigned to a particular class. . . . It is a standard language if it can be consistently applied to all individuals or social units. . . . Classifying observations into categories, ranking them, or counting instances serves to express observations in a language of measurement. . . . If these observations are expressed in a standard language, they are indeed comparable. (Przeworski & Teune, 1970, p. 93)

As discussed elsewhere (Pfau, 1980), such reasoning supports the choice of category systems for making comparisons. Category systems usually do provide specific, low-inference criteria for deciding whether certain classifications of behavior have occurred or not. The decisions made are expressed in a language of measurement, by using counting or other recording procedures based on time. The extent to which the language used is a standard language is indicated by the consistency or agreement obtained between different observers who use a particular category system.[2] That is, when different observers agree highly with respect to measurements made of the same phenomena, consistency in the use of the standard language is indicated. If this standard language is used to describe behaviors occurring in different nations or cultures, or at different times, the measurements made are comparable -- if one accepts the logic of Przeworski and Teune (see Pfau, 1980, for details).

Looked at in this way, <u>standardization</u> of behavioral measurements occurs when an observation instrument is used in such a way that (a) the behavioral categories of the instrument are used consistently by different persons to classify behaviors observed, and (b) measurements of the behaviors classified are made using the same metric so that scalar

identity is achieved across the occasions in which the instrument is used. This means that when the instrument is used to make measurements of behaviors occurring in different cultures or at different times, the measurements obtained should represent quantitatively identical scales (Poortinga, 1975). Differences in the measurements made should represent actual differences in the behavior observed, and not just differences in impressions made on individual observers, while equal scores should indicate equal amounts of the behavior observed.[3]

Past Usage of Category Systems

A number of researchers have used category systems to study behaviors occurring within single cultures as well as to compare behaviors occurring in different cultures and nations.

Cross-cultural studies in which these instruments were used have included studies of maternal and child behaviors in Japan and the U.S.A. (Caudill & Weinstein, 1969; Caudill & Schooler, 1973; Caudill & Frost, 1974), in Yugoslavia and the U.S.A. (Lewis & Ban, 1977), in different social classes and cultural groups in Israel (Greenbaum & Landau, 1977), as well as differences in infant separation protest in Guatemala and in the U.S.A. (Lester et al., 1974). A number of studies have also compared similarities and differences in classroom behaviors occurring in different countries including Britain and the U.S.A. (Birrell, 1974), Nepal and the U.S.A. (Pfau, 1977), and the Bahamas and the U.S.A. (Ray & Ray, 1976), as well as between schools of different cultural groups in the same country (Payne, 1970). Category systems have also been used to study the social patterns of urban pedestrians in Middle Eastern and Western countries (Berkowitz, 1971), and to study proxemic and other behaviors during conversations of Arabs and of Americans (Watson & Graves, 1966), and during

conversations in Germany, Italy, and the U.S.A. (Shuter, 1977).

Studies which used category systems to describe behaviors occurring within individual cultures have included ones which looked at maternal and child behaviors in Kenya, Guatemala, Botswana, Zambia, and Mexico (Leiderman, Tulkin, & Rosenfeld, 1977), numerous studies of teaching in different countries (e.g., Hilsum & Cane, 1971; Santhanam, 1972; Kim, 1972; Tisher, 1972; Flores, 1976; Rawakyaka, 1976; Hilsum & Strong, 1978; and Galton, 1979), and measurements of how members of an Amazon Indian group spent their time during the day (Johnson, 1975).

Variations of category systems have also been used in several significant cross-cultural studies. The most extensive and influential of these was the "Six Cultures Study", in which child-rearing and child behaviors in different cultures were described and compared (Whiting & Whiting, 1975). The approach used included having observers write extensive running accounts (called protocols) of behaviors which occurred and their contexts, rewriting the accounts in more and clearer detail as soon as possible, and then later coding the descriptions using a number of behavioral categories. Similar approaches, in which the running accounts have sometimes been called "specimen records", have included studies of behaviors occurring in an American and an English town (Barker & Barker, 1963, 1978; Schoggen, Barker, & Barker, 1963, 1978), studies of child behaviors in Japan and the U.S.A. (Caudill & Schooler, 1973), peer interaction in Kenya (Lubin & Whiting, 1978), and studies of Bengali infant and maternal behaviors (Graves, 1978).[4]

Another variation of category systems, the checklist, was used to study and compare science teaching in Britain and in Canada (Hacker, Hawkes, & Heffernan, 1979).

Perhaps a major accomplishment of these pioneering studies has been to demonstrate the range of theoretical and heuristic concerns, and the diversity of cultural and behavioral situations to which the systematic study of

behavior is applicable and fruitful. They have both stimulated thinking about what can be done, and provided a broad basis upon which future work and thinking can build.

For the most part, however, these studies represent only an incomplete beginning to standardizing behavioral measurements across cultures, for reasons to be now discussed.

Limitations of Past Category System Usage

A researcher studying behaviors occurring in different cultures, or in the same culture at different times, may wish to compare existing data gathered by others who used a particular category system. Or, he or she may conduct or coordinate studies aimed at describing behaviors in the cultures of interest, and then compare the resultant descriptions. A major consideration of researchers wishing to compare such data should be the "inter-investigation reliability" of the measurements used.

Inter-investigation reliability refers to the similarity of usage of the same instrument during different investigations (Rosenshine & Furst, 1973, p. 168). For our purposes, this concept is equivalent to "standardizing category system usage." That is, it refers to the question of whether the behavioral categories of an instrument are used the same way by different people to classify observations, and whether the same metric is used to do this so that measurements made have scalar identity.

Table 1 indicates some different ways of trying to ensure that standardization of category system and other observation instrument usage is achieved. These ways range from what this writer and others consider to be a "rigorous approach" (Longabaugh, 1980, pp. 104 & 106), to much more questionable approaches for measuring and comparing behaviors.

Using these approaches to classify the cross-cultural studies of behavior mentioned before provides the results

Table 1

Examples of Techniques for Helping Achieve
Standardization of Category System Usage
in Different Cultures

A. A Rigorous Approach[a]

Observers from Culture A and Culture(s) B (C,D,...) observe behaviors in Culture A using a category system and reach high agreement among measurements made. These same observers, without prior discussion, observe behaviors in Culture(s) B (C,D,...) using the category system and again reach high agreement among measurements made.

Such agreement indicates that a standard language and measurement procedures have been used in these cultures. Measurements of behaviors made by these observers using the category system are compared between the cultures.

B. Semi-Rigorous Approaches

Approach B_1: Observers from Culture A observe behaviors in Culture A using a category system and reach high levels of agreement among measurements made. One or more of these observers goes to Culture(s) B (C,D,...) and either makes observations directly using the category system or trains persons from Culture(s) B (C,D,...) to use the category system in that culture until their measurements agree highly with those made by the Culture A observer(s). Measurements made in Cultures A, B, ... using the category system are then compared.

Approach B_2: Same as B_1 except that persons from Culture(s) B (C,D,...) are observed in Culture A by observers from Culture A.

C. Approaches of Questionable Rigor

Approach C_1: An observer or observers attempt to use a category system the same way as others have used it, by reading descriptive materials or by learning to use it from a previous investigator. Formal checks of agreement are not made to determine similarity of usage with respect to other investigators, although accuracy checks may be made among observers using the instrument in the new study or sub-study. Measurements are made using the category system and compared with measurements made by other investigators who used the instrument in other cultures.

Table 1 (Continued)

Approach C_2: Observers reach understandings among themselves about how a category system is to be used. These observers then use the instrument in Cultures A, B, (C,D,...) and compare measurements made in these cultures. Formal checks of agreement are not made, however, to determine if these observers agree highly among themselves when observing and describing the same events.

Approach C_3: An observer uses the same category system to make observations in two or more cultures. Formal checks of agreement and stability of usage are not made, however.

D. Highly Questionable Approach

Measurements are compared of behaviors having the same general label, but which were measured using different instruments used by different investigators. No formal checks of agreement are made between measurements made using the different instruments when observing the same events.

---

[a]In this and the other approaches, standardization of usage also requires that other factors which may affect the accuracy of measurements made, such as "observer drift" and "time unit distortion", be controlled and eliminated (see Pfau, 1978, 1981, and Longabaugh, 1980, for details). For example, in order to avoid consensual observer drift, the observers observing in Culture(s) B (C,D,...) should not discuss observations made in that culture before formal checks of agreement are made between measurements made in that culture.

shown in Table 2. As can be seen, nearly all of the studies reviewed have used "semi-rigorous" to "questionable" approaches to standardize the measurements compared.

The main problem of the "semi-rigorous" approaches is that they did not ensure that observer drift in standardized usage did not occur when the observation instrument was used in another culture (Kazdin, 1977; Longabaugh, 1980, pp. 107-109). That is, these studies assumed that an observer or observers who used an observation instrument in standard ways in one culture, transferred standardized usage of the instrument to other cultures in which measurements were made. This is an untested assumption of these studies -- and can be viewed as a limitation of them and of "semi-rigorous" approaches in general.

The procedures used in the more "questionable" studies, besides not controlling for observer drift in usage, led to comparisons being made of data whose scalar identity is unknown and open to serious question -- and this is considered a serious limitation of these studies.[5]

Perhaps an even greater limitation, however, is that almost none of these studies has established a sufficient basis so that future researchers who may wish to gather and compare data with these past studies can ensure that their use of the "same" instrument is indeed the same. That is, almost none of these studies has provided enough information to permit standardization of observation instrument usage to be achieved between these past and future studies. This means that the scalar identity of past and future measurements cannot be estimated, and that comparisons of measurements made in the future with those made in these past studies will be hazardous.

As a result of this last limitation, the cross-cultural study of behavior is less cumulative than it might otherwise be. How this can be avoided in the future is indicated in the next section.

Table 2

Classification of Cross-Cultural Behavioral Studies with Respect to the Rigor of Standardization Procedures Used

A. Rigorous Approach Used

(no studies were reviewed which used such an approach)

B. Semi-Rigorous Approach

Approach B_1: Caudill and Weinstein (1969)
Payne (1970)
Berkowitz (1971)
Caudill and Frost (1974)
Lester et al. (1974)
Ray and Ray (1976)
Pfau (1977)
Shuter (1977)

Approach B_2: Watson and Graves (1966)

C. Approaches of Questionable Rigor

Approach C_1: Tisher (1970)
Hacker, Hawkes, and Heffernan (1979)[a]

Approach C_2: Barker and Barker (1963, 1978)
Schoggen, Barker, and Barker (1963, 1978)
Caudill and Schooler (1973)[b]
Whiting and Whiting (1975)

Approach C_3: Birrell (1974)

D. Highly Questionable Approach

Konner (1977, pp. 294-295)[c]

[a]This may in fact represent a B_1 approach if videotapes from the British investigation were used during observer agreement checks.

[b]It does not appear that formal checks were made of agreement between the "running accounts" made by the observers in this study. If such checks were made and were high, this would be a B_1 approach.

[c]Konner plainly admits that the Western data compared is "makeshift" (1977, p. 295).

A Suggested Approach

Techniques for helping overcome the problem of standardizing measurements made using category systems have already been indicated in Table 1. That is, one of the more rigorous approaches described in that table can be used to help ensure that observers are making standardized measurements in different locations and cultures at approximately the same time. However, the procedures indicated do not help to ensure that observers in the future will use an observation instrument as it was used in the past. This is so because even expert observers may modify their use of an instrument over time. (They may also die). Such "observer drift" in usage over time (more technically called "instrument decay") needs to be controlled if standardized measurements are to be made by observers at different times, either in the same or in future studies.

A way of doing this seems to be the following:[6]

1. Preserve Samples of Behavior Observed

Instrument developers and users should preserve samples of behavior observed on movie film, audio-visual tape recordings, audio recordings, or photographs -- the exact media used being dependent upon the types of behavior being measured.[7] Relevant contextual information which is not evident from the preserved behavioral records should also be described in sufficient detail so that future observers will have enough information to accurately code the behaviors preserved.

2. Make Standard Codings of These Preserved Behaviors

The preserved samples of behavior should then be coded by an "expert" observer, or by a typical group of observers used in the study of concern (whose codings are then averaged). These codings will constitute a set of

preestablished standards against which measurements made by others coding the preserved behavioral records can be compared.

3. Prepare Instrument Description and Training Materials

Sufficient information about the instrument used, including other preserved behavior samples and associated standard codings should be prepared so that future users can train themselves and others to use the instrument the same way it was used in the past ("Where do...," 1965; Thiagarajan, 1973). These materials should then be made available for future use by others. (See Herbert and Attridge, 1975, for suggestions about what to include in such training materials).

4. Future Users Should Conduct Standardization Checks

The materials developed would then be used by other investigators to train themselves or other observers to use the instrument. When usage is felt to be accurate, standardization checks would then be conducted by having observers code samples of the preserved behavioral records and checking their measurements against the preestablished standards prepared during step 2 described above. If high agreement is reached with these standard codings, then standardization of usage by the "new" observers is indicated. If high agreement is not reached, then a need for further training is indicated, as are afterwards, additional accuracy checks using previously unseen preserved behavioral records and associated standard codings.

5. Future Users Repeat Procedures 1 to 3

It is suggested that future users and adapters of the observation instrument also follow procedures 1, 2, and 3 described above to permit others to determine how the instrument was used by these researchers in their studies.

These procedures if followed and coupled with the techniques outlined in Table 1 can apparently be used to

overcome most if not all of the limitations of previous studies which were indicated before. For example, these procedures could be used in conjunction with the "Rigorous Approach" described in Table 1 by having observers from Culture A and Culture(s) B (C,D,...) receive training until their measurements of the preserved behavioral recordings agree highly with the preestablished standard codings of those recordings originally prepared. These observers would then proceed with the other observations in Culture A and Culture(s) B (C,D,...) outlined in Table 1.

Some Needed Research

Information helpful to persons wishing to standardize behavioral measurements across cultures and time would be provided by research which answered the following questions:

1. Can an observer who uses an observation instrument in a standard way in one culture, transfer standardized usage of the instrument to another culture (or language) when he or she observes behaviors in the second culture using the instrument or when he or she trains others in the second culture to use the instrument? An assumption of the widely used "Semi-Rigorous Approach: B_1" described in Table 1 is that, yes, such transferability of standardized usage can and does occur. As indicated before, this assumption is yet to be tested.

2. What technique(s) should investigators use to achieve standardization of observation instrument usage across cultures? The "Rigorous Approach" mentioned in Table 1 represents one way. Is this approach sufficient? Is it too rigorous? Are other approaches more practical and satisfactory?

3. Are the "weighting" procedures used by Caudill to help standardize measurements made by different observers in Japan and in the U.S.A. a promising approach to use when differences in instrument usage occur across cultures? (See Caudill and Weinstein, 1969, pp. 24-25, and Caudill and Frost, 1973, p. 7, for details of these procedures).

4. Does the two-step procedure involved in measuring behaviors by first writing specimen records/protocols and then coding these written descriptions, result in measurements which are accurate enough (i.e., are standardized enough) to compare across cultures? There are some indications that such may not be the case (Spain and Hollenbeck, 1975; Levine, 1977). The extent to which this procedure can be standardized across cultures needs to be studied more, given the fairly widespread use of such descriptions for comparative purposes.

FOOTNOTES

[1]For example, measurements can be made using these instruments by counting a behavior each time it occurs, by classifying and recording those classifications of behavior observed at very frequent intervals (e.g., every 3 or 5 seconds), by making classifications and recordings at longer but quite precise intervals of time (e.g., exactly every 5 minutes), by recording the duration of behaviors by using a stopwatch, or by noting the exact time behaviors of interest begin and end (Jackson, Della-Piana, & Sloane, 1975, pp. 33-34)

[2]Observers using category systems are ideally trained until their recordings agree highly with those made by experts in the use of the instrument. This high agreement helps ensure that observations are made in a standardized way. Also, the fact that high agreement can be reached indicates that the process of measurement is unambiguous, and that a "standard language" is being applied by the different observers to describe events observed (Pfau, 1980, p. 405).

[3]Such scalar identity is considered necessary if the scores of culturally different groups are to be compared, according to Davidson (1977, p. 50).

[4]Although the two-step (i.e., making a written description which is then coded) specimen record and protocol approaches provide a great deal of rich contextual information about behaviors observed, they seem to result in measurements which are not as accurate as those made by more directly coding behavior observed (Spain & Hollenbeck, 1975; Levine, 1977). Similarly, checklists represent another variation of category

systems. Checklists have relatively long time intervals during which the occurrence of behaviors are noted only once if they are observed. These instruments are less precise than category systems, are more prone to distorting estimates of the extent to which behaviors occurred, and do not lend themselves as well to the study of sequences of events (Dunkin & Biddle, 1974, p. 71). As a result, these variations do not seem to lend themselves as well to standardizing measurements across cultures, and are differentiated from category systems in this paper for that reason.

[5]This does not mean that scalar identity may not have been approximated in some cases, nor that a great deal of thought-provoking and useful data was not gathered by many of these studies. However, the degree to which measurements were made in a standardized manner in these studies is open to question.

[6]Readers are referred to Roebuck et al. (1974) for a related example of how standard usage of an observation instrument can be and has been maintained over an extended time period.

[7]For example, if locations of persons in a room are being studied, photographs may be sufficient. If an analysis of verbal behavior is being conducted, audio tape recordings may suffice. It should be noted though that specimen records, transcripts, or other kinds of narrative descriptions are not considered suitable for the preservation of realia as required in this step.

BIBLIOGRAPHY

Barker, R. G., & Barker, L. S. Social actions in the behavior streams of American and English children. In R. G. Barker (Ed.), The stream of behavior: Explorations of its structure & content. New York: Appleton-Century-Crofts, 1963.

Barker, R. G., & Barker, L. Social actions of American and English children and adults. In R. G. Barker & Associates, Habitats, environments, and human behavior. San Francisco: Jossey-Bass, 1978.

Berkowitz, W. R. A cross-national comparison of some social patterns of urban pedestrians. Journal of Cross-Cultural Psychology, 1971, 2, 129-144.

Birrell, H. V. A comparative study of the verbal behavior of teachers in open education classrooms in England and the United States (Doctoral dissertation, Syracuse University, 1974. Dissertation Abstracts International, 1976, 36, 6609A-6610A. (University Microfilms No. 76-7883)

Caudill, W., & Weinstein, H. Maternal care and infant behavior in Japan and America. Psychiatry, 1969, 32, 12-43.

Caudill, W. A., & Schooler, C. Child behavior and child rearing in Japan and the United States: An interim report. Journal of Nervous and Mental Disease, 1973, 157, 323-338.

Caudill, W., & Frost, L. A comparison of maternal and infant behavior in Japanese-American, American, and Japanese families. In W. P. Lebra (Ed.), Youth, socialization, and mental health: Vol III of mental health research in Asia and the Pacific. Honolulu: University Press of Hawaii, 1974.

Davidson, A. R. The etic-emic dilemma: Can methodology provide a solution in the absence of theory? In Y. P. Poortinga (Ed.), Basic problems in cross-cultural psychology. Amsterdam: Swets & Zeitlinger, 1977.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart, & Winston, 1974.

Ember, C. R. Feminine task assignment and the social behavior of boys. Ethos, 1973, 1, 424-439.

Flores, R. A description of the classroom verbal and non-verbal behavior of a selected group of junior high school science teachers in Puerto Rico (Doctoral dissertation, Temple University, 1976). Dissertation Abstracts International, 1976, 37, 2094A. (University Microfilms No. 76-22,040)

Galton, M. Systematic classroom observation: British research. Educational Research, 1979, 21, 109-115.

Graves, P. L. Infant behavior and maternal attitudes: Early sex differences in West Bengal, India. Journal of Cross-Cultural Psychology, 1978, 9, 45-60.

Greenbaum, C. W., & Landau, R. Mothers' speech and the early development of vocal behavior: Findings from a cross-cultural observation study in Israel. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Hacker, R. G., Hawkes, R. L., & Heffernan, M. K. A cross-cultural study of science classroom interactions. British Journal of Educational Psychology, 1979, 49, 51-59.

Herbert, J., & Attridge, C. A guide for developers and users of observation systems and manuals. American Educational Research Journal, 1975, 12, 1-20.

Hilsum, S., & Cane, B. S. The teacher's day. Windsor, Berks.: National Foundation for Educational Research in England and Wales, 1971.

Hilsum, S., & Strong, C. R. The secondary teacher's day. Slough: NFER Publishing, 1978.

Jackson, D. A., Della-Piana, G. M., & Sloane, H. N. How to establish a behavior observation system. Englewood Cliffs: Educational Technology Publications, 1975.

Johnson, A. Time allocation in a Machiguenga community. Ethnology, 1975, 14, 301-310.

Kazdin, A. E. Artifact, bias, and complexity of assessment: The ABCs of reliability. Journal of Applied Behavior Analysis, 1977, 10, 141-150.

Kim, C. S. A preliminary normative study on teaching behavior in elementary school. Seoul: College of Education, Seoul National University, 1972.

Konner, M. Infancy among the Kalahari Desert San. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Leiderman, P. H., Tulkin, S. R., & Rosenfeld, A. (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Lester, B. M., Kotelchuck, M., Spelke, E., Sellers, M. J., & Klein, R. E. Separation protest in Guatemalan infants: Cross-cultural and cognitive findings. Developmental Psychology, 1974, 10, 79-85.

Levine, A. J. Naturalistic observation: Validity of frequency data. Psychological Reports, 1977, 40, 1311-1338.

Lewis, M., & Ban, P. Variance and invariance in the mother-infant interaction: A cross-cultural study. In P. H. Leiderman, S. R. Tulkin, & A. Rosenfeld (Eds.), Culture and infancy: Variations in the human experience. New York: Academic Press, 1977.

Longabaugh, R. The systematic observation of behavior in naturalistic settings. In H. C. Triandis & J. W. Berry (Eds.), Handbook of cross-cultural psychology (Vol. 2). Boston: Allyn & Bacon, 1980.

Lubin, D., & Whiting, B. Techniques of persuasion in peer interaction. Paper presented at the annual meeting of the Society for Cross-Cultural Research, New Haven, February 1978.

Payne, J. I. Analysis of teacher-student classroom interaction in Amish and non-Amish schools using Flanders Interaction Analysis technique (Doctoral dissertation, Pennsylvania State University, 1970). Dissertation Abstracts International, 1971, 32, 251A. (University Microfilms No. 71-16,650)

Pfau, R. H. The cross-national comparison of classroom behaviors. Paper presented at the annual meeting of the Comparative and International Education Society, Toronto, February 1976. (ERIC Document Reproduction Service No. ED 120 260)

Pfau, R. H. A comparison of Nepalese and U.S. classroom behaviors. Paper presented at the annual meeting of the Comparative and International Education Society, New Orleans, February 1977a. (ERIC Document Reproduction Service No. ED 137 265)

Pfau, R. H. A cross-national comparison of classroom behaviors: Based upon a survey conducted within Nepal using the Flanders Interaction Analysis Categories and Caldwell's Activity Categories Instrument (Doctoral dissertation, University of Pittsburgh, 1977b). Dissertation Abstracts International, 1978, 38, 7034A. (University Microfilms No. 78-09,602)

Pfau, R. H. Factors affecting the validity of cross-national comparisons of classroom behaviors. Paper presented at the annual meeting of the Society for Cross-Cultural Research, New Haven, February 1978.

Pfau, R. H. The comparative study of classroom behaviors. Comparative Education Review, 1980, 24, 400-414.

Pfau, R. H. Sources of invalidity when comparing classroom behaviors across cultures and nations. Paper to be presented at the annual meeting of the Comparative and International Education Society, Tallahassee, March 1981.

Poortinga, Y. H. Some implications of three different approaches to intercultural comparison. In J. W. Berry & W. J. Lonner (Eds.), Applied cross-cultural psychology. Amsterdam: Swets & Zeitlinger, 1975.

Przeworski, A., & Teune, H. The logic of comparative social inquiry. New York: John Wiley, 1970.

Ray, R. D., & Ray, M. R. A systems approach to behavior II: The ecological description and analysis of human behavior dynamics. Psychological Record, 1976, 26, 147-180.

Roebuck, F. N., Aspy, D. N., Sadler, L. L., & Willson, M. A. Maintaining reliability in a longitudinal study. Monroe, Louisiana: National Consortium for Humanizing Education, Northeast Louisana University, 1974. (ERIC Document Reproduction Service No. ED 106 730)

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In M. W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973.

Rwakyaka, P. N. Teacher-student interaction in ESL classes in Uganda (Doctoral dissertation, Columbia University Teachers College, 1976). Dissertation Abstracts International, 1977, 37, 7003A-7004A. (University Microfilms No 77-5757)

Santhanam, M. R. Some explorative studies in India in classroom interaction analysis. Classroom Interaction Newsletter, 1972, 7, 14-22.

Schoggen, M., Barker, L. S., & Barker, R. G. Structure of the behavior of American and English children. In R. G. Barker (Ed.), The stream of behavior: Explorations of its structure & content. New York: Appleton-Century-Crofts, 1963.

Schoggen, M. F., Barker, L. S., & Barker, R. G. Behavior episodes of American and English children. In R. G. Barker & Associates, Habitats, environments, and human behavior. San Francisco: Jossey-Bass, 1978.

Shuter, R. A field study of nonverbal communication in Germany, Italy, and the United States. Communication Monographs, 1977, 44, 298-305.

Spain, D. H., & Hollenbeck, A. R. Coding non-Western behavior stream data: A problem of reliability. Journal of Social Psychology, 1975, 97, 137-138.

Thiagarajan, S. Instructional systems for interactional systems. Classroom Interaction Newsletter, 1973, 9, 13-22.

Tisher, R. P. A study of verbal interaction in science classes and its association with pupils' understanding in science. St. Lucia, Queensland: University of Queensland Press, 1970.

Tisher, R. P. Studies of behaviour in Australian classrooms. Classroom Interaction Newsletter, 1972, 7, 49-57.

Watson, O. M., & Graves, T. D. Quantitative research in proxemic behavior. American Anthropologist, 1966, 68, 971-985.

Where do we go from here? Classroom Interaction Newsletter, 1965, 1, 5.

Whiting, B. B., & Whiting, J. W. M. Children of six cultures: A psycho-cultural analysis. Cambridge, Mass.: Harvard University, 1975